Effects of Household Income on Graduation Level Across US Counties

Libby Boss, Graham Lewis, Karl Patram

Abstract

This paper analyzes the effect of household income on the graduation level on a county scale across the United States, following the hypothesis that household income should have a positive impact on graduation rates, measured in the percentage of people ages 18 to 24 with at least a high school degree as the dependent variable. Data from the 2010 US Census is used to estimate simple and multiple regression models. Other independent variables used are teen birth rate, average household size, poverty level, percent of people 25 and older with at least a high school degree, urban/rural location type, logarithm of household income, and percent of people ages 45 to 64 with at least a high school degree to take into account family history and regional differences and reduce omitted variable bias. The simple regression shows a positive relation between graduation rate and household income. The multiple regressions show that most independent variables have the hypothesized relation with the dependent variable, but fail to provide conclusive evidence about the hypothesis.

I. Introduction

"The source of America's prosperity has never been merely how ably we accumulate wealth, but how well we educate our people. This has never been more true than it is today. In a 21st-century world where jobs can be shipped wherever there's an Internet connection, where a child born in Dallas is now competing with a child in New Delhi, where your best job qualification is not what you do, but what you know -- education is no longer just a pathway to opportunity and success, it's a prerequisite for success."
-President Obama during speech to US Hispanic Chamber of Commerce 2009

The high school graduation rate is an important metric in the analysis of the U.S educational system. Research has found that graduating high school is a key indicator of the type of wages a worker will earn over the course of their lives (Jaeger and Page 1996). Even though a high school diploma has become commonplace compared to the workforce of 100 years ago, the impact of education can be felt across the economy (Goldin and Katz 2008). When considering the benefits of increases in education and cognitive levels have on countries (Shanushek and Woessmann 2008), the high school graduation rate should not be taken for granted. Currently, the United States ranks below the OECD average for secondary school graduation rates**. In the 1960s, the U.S. had the highest high school graduation rate in the world. However, the rate stagnated as other countries saw massive increases in their graduation rates Murnane (2013).

We expect the mean household income level for a county to be positively related to the percentage of the population of 18 to 24 years olds with a high school diploma or equivalent in that county. If household income has a very strong link to graduation rates, then the productivity of the workforce can be greatly increased by increasing household income. It would also mean that both economic mobility and inequality could be targeted by policies designed to lessen the gap between the education received by different socioeconomic levels.

We also are concerned with other non-monetary contributions of the society to education, such as if the parents are also educated, and proximity to cities. We assume that school quality within a county is relatively homogeneous, even though there will be differences between high schools (not taking into account the effects of local, district-only funding).

II. Literature Review

Stark, Noel, and McFarland (2015) analyzed key factors and statistics in high school dropouts. The research showed that in 2012, students living in low-income households had a 5.9% event dropout rate compared to the 1.6% event dropout rate enjoyed by students from high-income backgrounds. A 4.6% difference in the event dropout rate between low and high-income students showed a factor in the hindrance of economic mobility. In the paper, the term low-income is used to describe families whose incomes are in the bottom 20%. It follows that high income are families whose income is in the top 20%, leaving middle income to fill up the remaining 60%. From 1975, a decrease in the high school dropout rate is seen, followed by an increase in the early 1990s and then a decrease until a period of stagnation until 2012. The period from 1975 until 2012 saw a decrease in the dropout rates of high schoolers across all socioeconomic levels; however, the inequality still persisted.

Duncan, Kalili and Ziol-Guest (2017) looked at the demographics underneath the income-based gap in schooling. The demographics included family income, mother's education, family size, two-parent family structure, and age of mother at birth. They used data spanning 31 cohorts born between 1954 and 1985 from the Panel Study of Income Dynamics (PSID), finding that over the period of time they observed, the income gap grew between bottom quintile and top quintile income families. However, these were not the only demographic changes occurring during this time frame. For instance, the two-parent family structure rates decreased for low income families, while mother's age increased for higher income families. Duncan, Kalil and Ziol-Guest found that family income and maternal education were the largest indicators of children's education attainment. Thus, in addition to the simple regression of graduation rates on household income, it is important to do a multiple regression that controls for some of these other demographic factors.

Card and Krueger (2000) studied the effects of school quality on the rate of return to education, finding that school quality, measured in terms of such variables as student/teacher ratio, length of term, and relative teacher pay, is unsurprisingly positively related with the rate of return, measured in the logarithm of weekly earnings. We believe that a higher returns to education would incentivize individuals to complete more years of schooling, implying higher rates of high school graduation; therefore, the same independent variables should have the same effect on rates of high school graduation as rate of return to education. It is important to note that the Card and Krueger analysis discredits parental education and household incomes as having an effect on the rate of return to education, and specifically eliminates variables not intrinsic to the schools themselves, purposefully controlling for the effects of differing locations and family backgrounds; we will attempt to do the opposite, considering geographic and income differences as important factors in school quality, and including them explicitly as independent variables, to better focus on internal differences across the United States.

Kearney and Levine (2016) investigated whether locations burdened by high income inequality had lower rate of high school graduation among individuals who had low socioeconomic status. The research looked at the perceived return on investment by students on continuing their high school graduation, showing that students of low socioeconomic status perceived dropping out as more beneficial more often than their more wealthy counterparts. The research found that males who fit this category were much more likely to drop out if they lived in areas of high income inequality. The study also found that girls of low socioeconomic status living in such areas were much more likely to become unmarried, young mothers. The findings gave reasoning behind the perpetuation of economic inequality in that students were shown to respond from massive gaps between them and the middle class by dropping out.

This paper aims to determine the impact household income has on graduation rates by specifically taking into account variation on a county level across the United States, unlike previous papers that look at overall trends. We vary the size of the geography covered in the data, using the entire United States and looking at a county level rather than individual families. This paper uses more recent data than the previous literature, albeit for only one year (2010).

III. Data

The following variables were used for the analysis and will hereafter be referred to by their abbreviations. All data is taken at the county level.

Table 1. Variables

Variable (Units)	Abbreviation	Type	Year	Source
Percent of 18-24 year olds with a high school degree or equivalent (%)	HS1824	Dependent	2010	U.S. Census Bureau
Avg. Household Income, yearly ($)	HHIncMean	Independent	2010	U.S. Census Bureau
Average Household Size (people)	AHHSize	Independent	2010	U.S. Census Bureau
Percent of county that lives in an urban area (%)	Urban	Independent	2010	U.S. Census Bureau
Percent of county below poverty line (%)	PovertyLevel	Independent	2010	U.S. Census Bureau
Percent of people ages 25+ with at least high school degree (%)	HSorAbove25	Independent	2010	U.S. Census Bureau
Teen births per 1000 teen females aged 15-19 (‰=0.1%)	TeenBirthRate	Independent	2010	U.S. Data.GOV
Logarithm of average household income, yearly ($)	lHHIncMean	Independent	2010*	(*taken by the natural log of HHIncMean)
Percent of people ages 45-64 with at least HS degree (%)	ParEduc	Independent	2010	U.S. Census Bureau

A brief discussion of each of the variables follows, as well as their summary statistics.

<u>Dependent variable</u>

HS1824

The dependent variable in all regressions is the percentage of the population of 18 to 24 year-olds with a high school degree or equivalent, used as a proxy for graduation rates. The use of the percentage, rather than simply the total number of people, is to adjust for the differing total population sizes.

Independent Variables

HHIncMean

The census gives household income in both mean and median form; we are using the mean rather than the median to recognize the effects of outliers. This variable is used as an indicator of the county's wealth level and is our primary interest; it is used in the simple as well as the multiple regressions. In counties that have a high mean household income, the budgets for schools should be larger therefore directly contributing to higher graduation rates.

HSorAbove25

A higher value represents a more educated population, meaning students who grew up in these counties were surrounded by adults who had already judged it worthy of their time to graduate high school, raising the perceived benefits of education and incentivizing high school graduation. This variable is used as a proxy for the education of the parents.

Urban

The US Department of Education reported that high school students in rural areas graduate at a higher rate than those in urban areas (Marcus 2018). One of the reasons discussed is that there are usually more distractions to students' schooling in urban areas, such as crime. Because of this statistic, we expect that those living in urban areas will graduate at a lower rate in this model.

PovertyLevel

Households that are below the poverty level are expected to have lower rates of graduation for the children. Previous literature has suggested that this may be due to many parents in this category not having enough time to spend with their children because they may work extra jobs to get the income they have. Children living in these households often have to work to help contribute to the household income from an earlier age, with some choosing to work instead of completing high school. Higher poverty affects students' perceived return on investment for graduation (Kearney and Levine 2016).

AvgHHSize

The number of children a set of parents have can decrease an individual child's education attainment because a parent has less time available to help each child with school work or other activities. A large household also creates the possibility of older siblings dropping out to support a large family. The counties with higher averages are expected to have lower rates of graduation in the 18-24 age range.

TeenBirthRate

When a teen becomes a mother they have to provide for a family at a young age; when they are in a lower income area, they drop out at higher rates than those in higher income areas to be able to get jobs to support themselves. We expect that the higher this rate is, the lower the graduation rate for that county. Girls in low income areas are much more likely to drop out than their richer counterparts (Kearney and Levine 2016). Although the data in the regression is in units of permille, we will generally interpret it with the equivalent unit of 0.1%.

lHHIncMean

This variable replaced HHIncMean in later models to better account for the shape of the data (see Appendix A); income tends to be positively skewed.

ParEduc

This variable replaced HSorAbove25 to better account for possible ages of the parents. In a county with more parents being high school graduates, it would make sense for that county to experience higher levels of high school graduation rates because graduation is seen as an expectation instead of a rare occurrence.

Table 2. Summary Statistics

Variable (abbrev.)	Observations	Mean	Standard Deviation	Minimum	Maximum
HS1824	3,135	78.81%	9.81%	19.5%	100%
HHIncMean	3,135	$56,618.57	$13,794.12	$28,594	$137,811
HSorAbove25	3,135	35.64%	6.92%	8.2%	54.6%
Urban	3,135	40.53%	31.52%	0%	100%
PovertyLevel	3,135	20.90%	9.82%	0%	63.28%
AvgHHSize	3,135	2.51	0.25	1.76	4.47
TeenBirthRate	3,135	4.057%	1.858%	0.410%	12.493%
lHHIncMean	3,135	10.92	0.2183	10.26	11.83
ParEduc	3,135	86.23%	7.698%	40.8%	100%

We perform a simple regression of HS1824 on HHIncMean, as well as three multiple regression models using various independent variables. We first check if these regressions satisfy the Gauss-Markov assumptions.

Linear in Parameters

We assume that in the population, the dependent variable can be written as a linear combination of the independent variables, plus an error term; the scatter graphs between each of the independent variables and the dependent variable (included in Appendix A) appear to come from linear relationships, satisfying our assumption.

Random Sampling

All data is taken from the US Census or data.Gov, two government agencies who gather data on known households. Sampling in this situation is random as there are no groups specifically targeted when collecting data. Every known house is targeted for data collection, making these sources very reliable.

No Perfect Collinearity

It is evident that none of the independent variables are constant in the sample. Second, it is reasonable to assume that there are no direct linear relationships between any combination of them, since in the sample, no correlation coefficient between any two independent variables is not close to 1 (see Appendix B). This satisfies the assumption of no perfect collinearity.

Zero Conditional Mean

We assume that the independent variables contain no information about the size of the error; zero conditional mean is not likely to be satisfied in the simple regression due to omitted variable bias, but is more likely to be satisfied in the multiple regression. However, there will likely always be information about the unobserved factors not accounted for in the independent variables. Formally, we assume that for our *k* variables, $\mathrm{E}(u|x_1, \ldots, x_k) = 0$.

Homoskedasticity

We assume that the independent variables contain no information about the variance of the error; formally, $\mathrm{Var}(u|x_1, \ldots, x_k) = \sigma^2$.

Because the data comes from the census and includes over 3,000 data points, we may safely assume that the regression does not suffer from the problem of micronumerosity caused by small samples with little variation in the independent variables. Moreover, given the “low” correlation coefficients between the independent variables in Appendix B, we may assume that the regression does not suffer from multicollinearity.

III. Results

Simple Linear Regression and Multiple Linear Regression 1

Table 3. Estimated Parameters and Standard Errors, SLR and MLR1

Dependent Variable HS1824		
Independent Variables	SLR	MLR 1
HHIncMean	.0002069 (0.0000122)***	-.0000593 (0.0000173)***
Urban		.0760405 (0.005643)***
HSorAbove25		.209553 (0.03314)***
PovertyLevel		-.0765522 (0.02399)***
AvgHHSize		-3.08814 (0.7241)***
TeenBirthRate		-.1477874 (0.01311)***
Intercept	67.09139 (0.7087)***	77.0277 (3.7938)***
No. of Obs.	3,135	3,135
R-square	0.0843	0.2760

Estimated Value (Std. Error) *Significant at 10%, **5%, ***1%

Simple Linear Regression: $y = \widehat{\beta}_0 + \widehat{\beta}_1 x + u$

HS1824 = 67.0914 + .0002(HHIncMean) + *u*

The simple regression shows a positive relation between HS1824 and HHIncMean ($\beta_1 = .0002$), as predicted. We note that measuring HHIncMean in thousands of dollars, rather than dollars, changes the slope parameter to $\beta_1' = .2$, a result that is more useful for interpretation: the regression predicts that for each increase of a thousand dollars in mean household, there is a 0.2% point increase in graduation rates. Equivalently, an increase of five thousand dollars in mean household income can be expected to produce a 1% point increase in graduation rates, an economically significant change. However, the very low R-squared value of 0.085 evidences that very little of the variation in graduation rates is captured by this regression, and prompts us to include other relevant variables in hopes of capturing this variation.

Multiple Linear Regression 1: $y = \widehat{\beta}_0 + \widehat{\beta}_1 x_1 + \widehat{\beta}_2 x_2 + \widehat{\beta}_3 x_3 + \widehat{\beta}_4 x_4 + \widehat{\beta}_5 x_5 + \widehat{\beta}_6 x_6 + u$

HS1824 = 77.0277 - .00006(HHIncMean) + .0760(Urban) + .2096(HSorAbove25) - .0766(PovertyLevel) - 3.0881(AvgHHSize) - .1479(TeenBirthRate) + *u*

Whether each independent variable has a positive or negative effect on HS1824 is given by its sign in the equation. To better give a sense of the magnitudes of the effects of each of the independent variables, the model predicts that a 1% point increase in HS1824 can be caused by any of exactly one of the following changes, ceteris paribus: HHIncMean *decreases* by $16,667; Urban *increases* by 13.158 percentage points; HSorAbove25 *increases* by 4.77 percentage points; PovertyLevel *decreases* by 13.054 percentage points; AvgHHSize *decreases* by .32 people; TeenBirthRate *decreases* by 0.676 percentage points.

For all of the independent variables but two, the sign of the relationship between the dependent variable and each of the independent variables, given by the sign of their individual slope parameters, is as hypothesized. The exceptions are Urban and HHIncMean.

The relationship between HS1824 and Urban is positive, when we predicted it to be negative. Simply, our prediction could have been incorrect; there are factors of urban centers (such as access to technology, proximity to schools) that would positively influence graduation rates, and these positive factors must outweigh the negative ones.

Particularly surprising is the negative relation between HS1824 and HHIncMean. The variables themselves are positively correlated (for all of the other independent variables, the sign of the slope parameter matches the sign of the correlation with the dependent variable, as expected). Furthermore, HHIncMean showed a positive relationship in the simple regression.

This anomaly may be due to the interference of collinearity; that is, the true "positive" effect of HHIncMean is eaten away by the presence of related secondary independent variables that are now explicitly included in the regression, the result of a violation of the third Gauss-Markov assumption.

Looking at the correlation coefficients table, we note that all of the coefficients above .50 involve either PovertyLevel or HSorAbove25, making them prime suspects for collinearity. This makes intuitive sense, since PovertyLevel depends directly on HHIncMean, while HSorAbove25 is clearly related to HS1824, since members of the second category become members of the first with stunning ease (albeit requiring a little bit of time). These two variables were therefore eliminated in following regressions.

Multiple Linear Regressions 2 and 3

Table 4. Estimated Parameters and Standard Errors, MLR2 and MLR3

Dependent Variable HS1824		
Independent Variables	MLR 2	MLR 3
HHIncMean	-0.00000517 (0.00000154)	
Urban	0.07895 (0.005586)***	0.07876 (0.005676)***
AvgHHSize	-4.7499 (0.6983)***	-4.9805 (0.7059)***
TeenBirthRate	-0.2063 (0.01096)***	-0.2113 (0.01171)***
lHHIncMean		-0.05579 (1.0105)
ParEduc		-0.3756 (0.02292)
Intercept	96.2099 (1.6205)***	100.5635 (10.7916)***
No. of Obs.	3,135	3,135
R-squared	0.2612	0.2618

Estimated Value (Std. Error) *Significant at 10%, **5%, ***1%

MLR2: $y = \widehat{\beta}_0 + \widehat{\beta}_1 x_1 + \widehat{\beta}_2 x_2 + \widehat{\beta}_3 x_3 + \widehat{\beta}_4 x_4 + u$

HS1824 = 96.20995 - 5.17E-6(HHIncMean) + .078949(Urban) - 4.749923(AvgHHSize) - .2062998(TeenBirthRate) + u

MLR3: $y = \widehat{\beta}_0 + \widehat{\beta}_1 x_1 + \widehat{\beta}_2 x_2 + \widehat{\beta}_3 x_3 + \widehat{\beta}_4 x_4 + \widehat{\beta}_5 x_5 + u$

HS1824 = 100.5635 -.0557866(lHHIncMean) + .0787621(Urban) - 4.980473(AvgHHSize) - .211304(TeenBirthRate) - .037651(ParEduc)

The second MLR sought to improve the first by eliminating variables thought to be interfering due to collinearity. The estimated effect of HHIncMean was negative, contrary to the hypothesis; however, the estimated value was not significant, which we discuss in the following section.

The third MLR replaced HHIncMean with its lHHIncMean, and included the new variable ParEduc. Once again, the estimated effect of the primary variable, lHHIncMean was negative, but not significant. Similarly, the estimated slope parameter of ParEduc was negative, contrary to the hypothesis, but not significant.

The R-squared value in both regressions represented a decrease from the first multiple regression, which is unsurprising given that they included less variables. However, it also suggests that the variables specified in these models were no better in explaining the variation in high school graduation than those in the first.

Statistical Inference

The t-statistics, p-values, and 95% confidence intervals for the estimated parameters for each of the four models is included in Appendix C. For nearly all of the variables, the estimated parameters were significant at 1%, which suggests that the model is not overspecified. Most relevant to the hypothesis, we discuss the 95% confidence interval of the primary independent variable in each of the regressions (HHIncMean in SLR and MLR1, lHHIncMean in MLRs 2 and 3, both variables hereafter referred to as simply "income" and "log income"); we also look at ParEduc as the only other variable not significant at 1% in all regressions.

As noted before, the simple regression predicted a significantly positive effect of income on graduation rates, consistent with the correlation between the two variables. In saying "significantly positive," we mean that the 95% confidence interval contains only positive values, supporting the hypothesis that the effect of income on graduation rates is indeed positive. On the other hand, MLR1 predicted a significantly negative effect of income on graduation rates, strong evidence against the hypothesis.

Because of this unexpected outcome, later models sought to better specify variables and isolate the true effect of income. However, MLR2 predicted a negative effect of income on graduation rates, but not a significant one, evidenced by the fact that the confidence interval contains both negative and positive numbers. That is, although the model predicted, based on the data, a negative effect, it also concedes that it is plausible that the overall, true effect is positive. Whether the hypothesis is true or false is inconclusive. Similarly, MLR3 predicted a negative effect of log income on graduation rates, but not a significant one, leaving the hypothesis inconclusive.

Thus, the changes made in specifying MLRs 2 and 3 failed to provide statistical evidence to reject or support the hypothesis (the coefficient was not significant).

ParEduc was the only other variable not to be strongly significant in the regression(s) in which it was included; moreover, its slope in MLR3 was negative, contrary to the expected. One possible explanation is that the variable of parental education is irrelevant, which we attempt to refute in the following section. Another explanation is that the variable as defined (see Table 1) is not a good measurement for parents' education.

IV. Extensions

Given the strong significance of nearly all of the variables used, the presence of two non-significant variables in MLR3, lHHIncMean and ParEduc, opens the door for an F-test. As noted before, if these two variables were individually significant, because their slopes are negative, we would reject our hypothesis. However, since they are not individually significant, it is possible that they are irrelevant to the analysis, as Card and Krueger (2000) found in the case of returns to education. We thus run a test of joint significance on lHHIncMean and ParEduc, concluding that they are irrelevant if their slopes are both zero.

Using MLR3 as the unrestricted model, and the restricted model as given in Appendix D, we calculate the F-statistic for the hypothesis $H_0 : \beta_1 = 0 = \beta_5$ vs. $H_A :\ H_0\ \textit{not true}$, with 2 restrictions and 3,129 degrees of freedom, getting $F = \frac{(267988.867-222833.908)/2}{222833.908/3129} = 317$. Since the F-statistic is far greater than the 5% critical value of 3.00, we overwhelmingly reject the null hypothesis in favor of the alternate that the variables are jointly significant.

To conclude, neither lHHIncMean nor ParEduc is irrelevant for explaining graduation rates, but MLR3 fails to provide a significant estimate of whether their effects are positive or negative. Thus, it is probable that MLR3 is misspecified.

V. Conclusions

This paper aimed to show the effects of average household income on high school graduation rates per county across the United States with the hypothesis that there would be a positive correlation between average household income and high school graduation rates. The data showed that household income had a significant positive effect on high school graduation rates in a simple regression model, but a significant negative effect in a multiple regression model. Later multiple regression models, built on improving specification, showed a very weak negative effect that was not statistically significant. Considering the data collected, our hypothesis - high school graduation rates per county are positively correlated with median household income per county - can be neither rejected nor validated. Further, a joint test of significance showed that the logarithm of household income, along with parents' education, is significant.

Overall, the regressions explained very little of the variance of high school graduation rates, though most of the variables used were individually significant. With the collection of more data and the inclusion of more, well-specified variables, a superior explanation can be offered. Policies intended to raise graduation rates would need to consider variables that provide a higher explanation of the variance in high school graduation rates in order to ensure a variable with more explanatory power has not been unaccounted for.

Finally, the study was limited in its ability to use data regarding individual school levels of counties across the United States. A future project might benefit from incorporating school performance data across districts, rather than counties.

References

Card, David, and Alan B. Krueger. "Minimum Wages and Employment: A Case Study of the Fast-Food Industry in New Jersey and Pennsylvania: Reply." *The American Economic Review*, vol. 90, no. 5, 2000, pp. 1397–1420. *JSTOR*, www.jstor.org/stable/2677856.

Duncan, Greg J., et al. "Increasing Inequality in Parent Incomes and Children's Schooling." *Demography*, vol. 54, no. 5, 2017, pp. 1603–1626., doi:10.1007/s13524-017-0600-4.

Goldin, Claudia, and Lawrence Katz. "The Race between Education and Technology: The Evolution of U.S. Educational Wage Differentials, 1890 to 2005." 2007, doi:10.3386/w12984.

Hanushek, Eric A, and Ludger Woessmann. "The Role of Cognitive Skills in Economic Development." *Journal of Economic Literature*, vol. 46, no. 3, 2008, pp. 607–668., doi:10.1257/jel.46.3.607.

Jaeger, David A & Page, Marianne E, 1996. "Degrees Matter: New Evidence on Sheepskin Effects in the Returns to Education," The Review of Economics and Statistics, MIT Press, vol. 78(4), pages 733-740, November.

Marcus, Jon, and Matt Krupnick. "Who's Missing From America's Colleges? Rural High School Graduates." *NPR*, NPR, 15 Feb. 2018, www.npr.org/sections/ed/2018/02/15/581895659/whos-missing-from-america-s-colleges-rural-high-school-graduates.

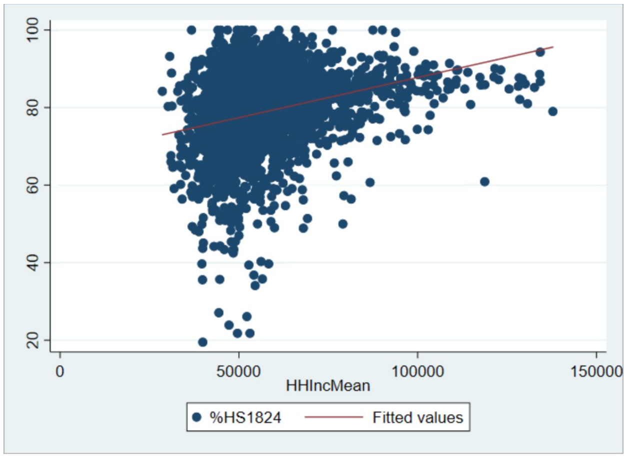
Melissa S. Kearney & Phillip B. Levine, 2016. "Income Inequality, Social Mobility, and the Decision to Drop Out of High School," Brookings Papers on Economic Activity, vol 2016(1), pages 333-396.

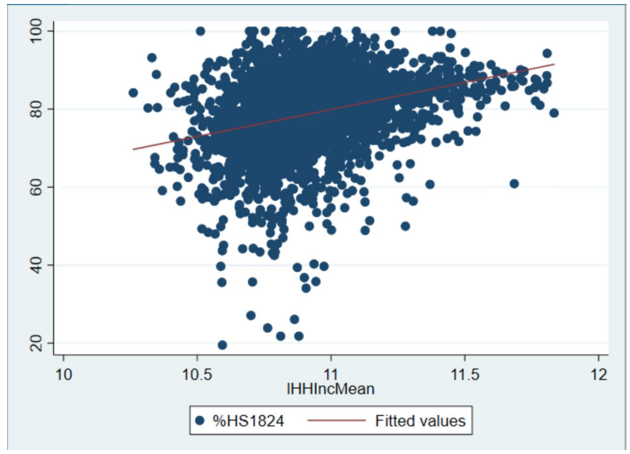
Richard J. Murnane, 2013. "U.S. High School Graduation Rates: Patterns and Explanations," Journal of Economic Literature, American Economic Association, vol. 51(2), pages 370-422, June.

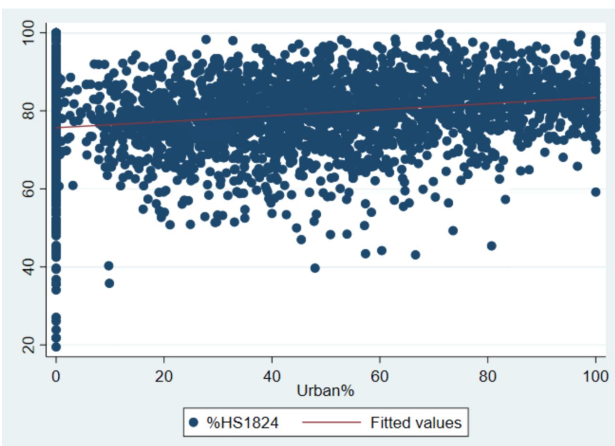
Stark, P., and Noel, A.M. (2015). Trends in High School Dropout and Completion Rates in the United States: 1972–2012 (NCES 2015- 015). U.S. Department of Education. Washington, DC: National Center for Education Statistics.

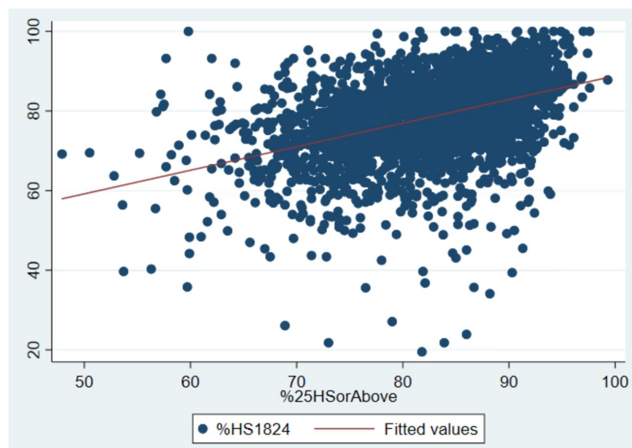
Appendix A. Scatter Graphs of Independent Variables

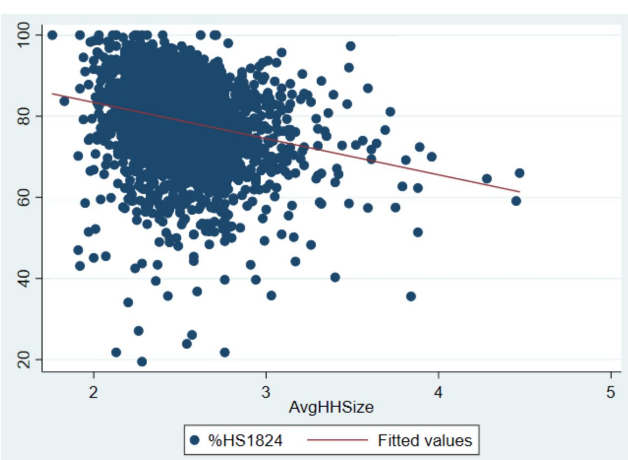
The scatter graphs between each of the independent variables and the dependent variables show a relationship that is believably linear.

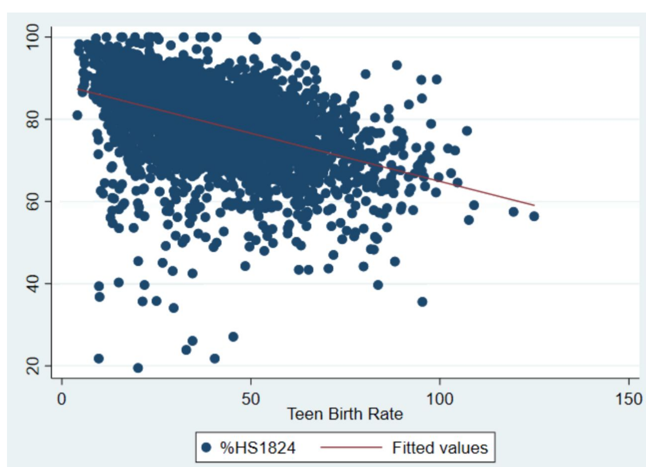


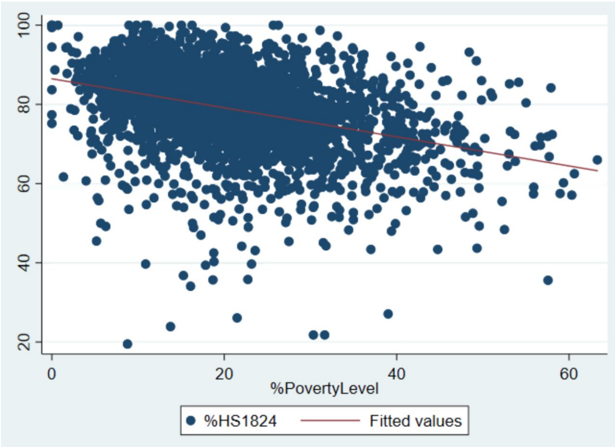


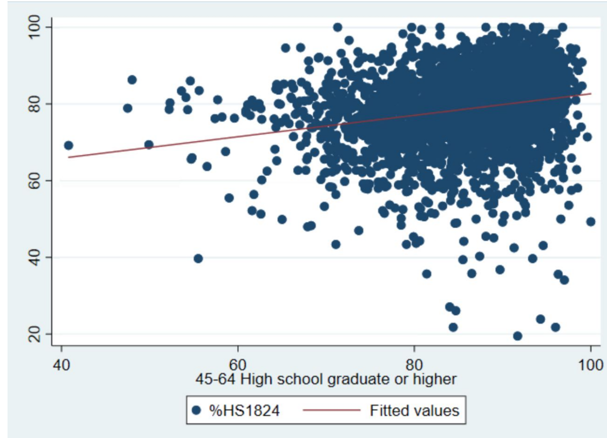









100
80
60
40
20
0
20
40
60
%PovertyLevel
%HS1824
Fitted values


100
80
60
40
20
40
60
80
100
45-64 High school graduate or higher
%HS1824
Fitted values

Appendix B. Correlation Coefficient Table

MLR 1

	HS1824	HHIncM~n	Urban	HSorA~25	Povert~l	AvgHHS~e	TeenBi~e
HS1824	1.0000						
HHIncMean	0.2908	1.0000					
Urban	0.2485	0.4696	1.0000				
HSorAbove25	0.4442	0.5287	0.2091	1.0000			
PovertyLevel	-0.3670	-0.6171	-0.1369	-0.6605	1.0000		
AvgHHSize	-0.2266	0.1102	0.1564	-0.3664	0.2074	1.0000	
TeenBirthR~e	-0.4429	-0.4924	-0.0440	-0.7176	0.6732	0.3706	1.0000

MLR 3

	HS1824	lHHInc~n	Urban	AvgHHS~e	TeenBi~e	ParEduc
HS1824	1.0000					
lHHIncMean	0.3083	1.0000				
Urban	0.2485	0.4816	1.0000			
AvgHHSize	-0.2266	0.0857	0.1564	1.0000		
TeenBirthR~e	-0.4429	-0.5190	-0.0440	0.3706	1.0000	
ParEduc	0.2194	0.3068	0.0580	-0.2970	-0.4927	1.0000

Appendix C. Stata Outputs

SLR

Source	SS	df	MS			
				Number of obs	=	3,135
				F(1, 3133)	=	289.48
Model	25532.77	1	25532.77	Prob > F	=	0.0000
Residual	276341.247	3,133	88.2033983	R-squared	=	0.0846
				Adj R-squared	=	0.0843
Total	301874.017	3,134	96.3222773	Root MSE	=	9.3917

HS1824	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
HHIncMean	.0002069	.0000122	17.01	0.000	.0001831	.0002308
_cons	67.09139	.708721	94.67	0.000	65.70178	68.48099

MLR1

Source	SS	df	MS			
				Number of obs	=	3,135
				F(6, 3128)	=	198.77
Model	83327.0911	6	13887.8485	Prob > F	=	0.0000
Residual	218546.926	3,128	69.8679431	R-squared	=	0.2760
				Adj R-squared	=	0.2746
Total	301874.017	3,134	96.3222773	Root MSE	=	8.3587

HS1824	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
HHIncMean	-.0000593	.0000173	-3.43	0.001	-.0000932	-.0000254
Urban	.0760405	.005643	13.48	0.000	.0649762	.0871048
HSorAbove25	.209553	.0331428	6.32	0.000	.1445691	.274537
PovertyLevel	-.0765522	.0239976	-3.19	0.001	-.1236049	-.0294996
AvgHHSize	-3.08814	.7240985	-4.26	0.000	-4.507897	-1.668384
TeenBirthR~e	-.1477874	.013109	-11.27	0.000	-.1734906	-.1220842
_cons	77.0277	3.793812	20.30	0.000	69.58909	84.46631

MLR 2

```
. regress HS1824 HHIncMean Urban AvgHHSize TeenBirthRate
```

Source	SS	df	MS			
				Number of obs	=	3,135
				F(4, 3130)	=	276.66
Model	78852.0446	4	19713.0111	Prob > F	=	0.0000
Residual	223021.972	3,130	71.2530263	R-squared	=	0.2612
				Adj R-squared	=	0.2603
Total	301874.017	3,134	96.3222773	Root MSE	=	8.4412

HS1824	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
HHIncMean	-5.17e-06	.0000154	-0.33	0.738	-.0000355	.0000251
Urban	.078949	.0055864	14.13	0.000	.0679956	.0899025
AvgHHSize	-4.749923	.6982779	-6.80	0.000	-6.119052	-3.380794
TeenBirthR~e	-.2062998	.0109592	-18.82	0.000	-.2277876	-.1848119
_cons	96.20995	1.620541	59.37	0.000	93.03252	99.38739

MLR 3

Source	SS	df	MS			
				Number of obs	=	3,135
				F(5, 3129)	=	221.97
Model	79040.1091	5	15808.0218	Prob > F	=	0.0000
Residual	222833.908	3,129	71.2156944	R-squared	=	0.2618
				Adj R-squared	=	0.2607
Total	301874.017	3,134	96.3222773	Root MSE	=	8.4389

HS1824	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
lHHIncMean	-.0557866	1.010503	-0.06	0.956	-2.037103	1.92553
Urban	.0787621	.0056755	13.88	0.000	.0676339	.0898902
AvgHHSize	-4.980473	.7058819	-7.06	0.000	-6.364512	-3.596435
TeenBirthR~e	-.211304	.0117108	-18.04	0.000	-.2342656	-.1883424
ParEduc	-.037651	.0229186	-1.64	0.101	-.0825881	.0072861
_cons	100.5635	10.79163	9.32	0.000	79.40412	121.7229

Appendix D. Restricted Model for F-Test

Source	SS	df	MS			
				Number of obs	=	3,135
				F(3, 3131)	=	368.95
Model	78844.0706	3	26281.3569	Prob > F	=	0.0000
Residual	223029.946	3,131	71.2328158	R-squared	=	0.2612
				Adj R-squared	=	0.2605
Total	301874.017	3,134	96.3222773	Root MSE	=	8.44

HS1824	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
Urban	.0780353	.0048725	16.02	0.000	.0684817	.087589
AvgHHSize	-4.82379	.6623529	-7.28	0.000	-6.12248	-3.5251
TeenBirthR~e	-.2041113	.0087912	-23.22	0.000	-.2213483	-.1868743
_cons	96.05137	1.549431	61.99	0.000	93.01336	99.08937